APPENDIX A

$\Delta = b^2 - 4ac$*

BY JEAN-PIERRE SERRE
Collège de France

The formula of the title is of course familiar; it is the *discriminant* of the quadratic polynomial $ax^2 + bx + c$.

The problem I want to discuss today is: Given an integer Δ, *what are the possible polynomials $ax^2 + bx + c$, with integer coefficients a, b, c, for which $b^2 - 4ac$ is equal to Δ?* Can we classify them?

This problem has a long history, going as far back as Gauss (circa 1800); it is not solved yet, but there have been quite exciting new results recently, as I hope to show you.

Notice first that there is an obvious necessary condition on Δ; namely Δ should be congruent to a square mod 4, i.e.,

$$\Delta \equiv 0, 1 \pmod 4.$$

Conversely, if this congruence holds, it is easy to find $a, b, c \in \mathbb{Z}$ with $\Delta = b^2 - 4ac$ (exercise). This settles the question of the *existence* of the solutions of our problem; it remains only (!) to classify them. For instance, are there some Δs for which there is a unique solution?

In this crude form, the answer is obviously "no." Indeed, the transformation $x \to x + 1$ leaves Δ invariant, but changes (a, b, c) to $(a, b + 2a, a + b + c)$. Thus, we should consider two quadratic polynomials as *equivalent* if

*Lecture organized jointly by the Singapore Mathematical Society and the Department of Mathematics, National University of Singapore, and delivered on 14 February 1985. Notes taken by Daniel E. Flath.

they differ by $x \to x + 1$, or more generally, by $x \to x + n$ ($n \in \mathbb{Z}$). But this is not enough: There are other possible transformations. To see them, it is better to use a homogeneous notation and to write our quadratic polynomials as $ax^2 + bxy + cy^2$. The transformation $x \to x + 1$ becomes $\begin{cases} x \to x + y \\ y \to y \end{cases}$, which we may write as a matrix $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Since now x and y play symmetric roles, we should introduce as well the matrix $T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, which corresponds to the transformation $\begin{cases} x \to x \\ y \to x + y \end{cases}$. And, since we can compose transformations, we should consider the group generated by S and T, which happens to be the group $\mathbf{SL}_2(\mathbb{Z})$ of two-by-two matrices $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, with integral coefficients and determinant 1.

Now our problem may be reformulated as follows:

Given an integer Δ, with $\Delta \equiv 0, 1 \pmod 4$, classify the $\mathbf{SL}_2(\mathbb{Z})$ equivalence classes of quadratic forms $ax^2 + bxy + cy^2$, with $a, b, c \in \mathbb{Z}$ and $b^2 - 4ac = \Delta$.

For the rest of this talk, we will consider *only the case where Δ is* < 0, i.e., equations $ax^2 + bx + c = 0$ with no real root. (The case of a positive Δ is equally interesting, but quite different, and there has been little progress on it since Gauss.) This restriction to negative Δs forces a and c to have the same sign. For convenience, we will always take them positive, and we will denote by $\underline{h}(\Delta)$ the number of such forms, modulo $\mathbf{SL}_2(\mathbb{Z})$ equivalence; we shall see below that this number is finite.

Consider a form $ax^2 + bxy + cy^2$, with $a, c > 0$, and $b^2 - 4ac = \Delta$, with $\Delta < 0$. We say that such a form is *almost reduced* if $a \le c$ and $|b| \le a$. *Any form can be transformed into an almost reduced one* by an element of $\mathbf{SL}_2(\mathbb{Z})$. Indeed, we can arrange that $a \le c$ by applying the transformation $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in case $c < a$ and we can ensure that $|b| \le a$ by applying some shift $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$, which leaves a invariant and replaces b by $b + 2an$. If this destroys the inequality $a \le c$, we apply again $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and so on. It is easily checked that this process comes to a stop after finitely many steps and gives an almost reduced form.

Theorem. The number of almost reduced forms with given discriminant $\Delta < 0$ is finite.

Proof. If $ax^2 + bxy + cy^2$ is almost reduced, we have

$$4a^2 \le 4ac = b^2 - \Delta \le a^2 - \Delta,$$

hence $3a^2 \le -\Delta$; this shows that a can take only finitely many values. The same is true for b since $|b| \le a$, and c is determined by a, b, and Δ. ∎

Corollary. $\underline{h}(\Delta)$ is finite.

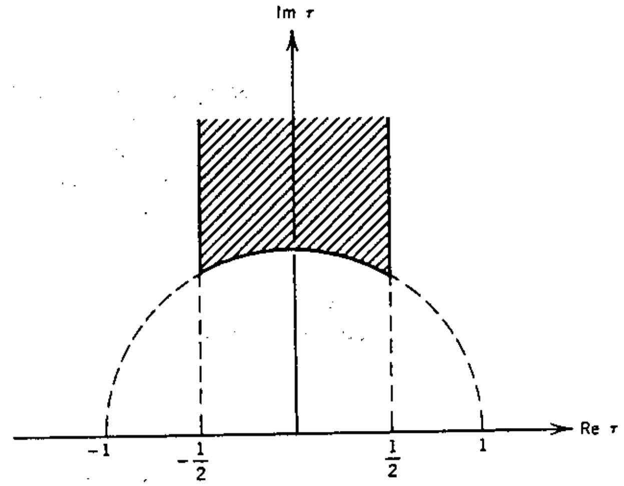

Figure A1

To go further, we need to investigate whether every $\mathbf{SL}_2(\mathbf{Z})$ equivalence class contains a *unique* almost reduced form. It turns out that this is nearly always true. I want to explain the exceptions by using a picture in the complex plane: Write $ax^2 + bxy + cy^2$ as $a(x + \tau y)(x + \bar{\tau} y)$ with some complex number τ. We may assume that $\operatorname{Im} \tau > 0$ since τ and $\bar{\tau}$ play symmetric roles. The condition $|b| \le a$ is equivalent to $|\tau + \bar{\tau}| \le 1$, that is $|\operatorname{Re} \tau| \le \frac{1}{2}$. The condition $a \le c$ translates to $\tau\bar{\tau} \ge 1$, that is $|\tau| \ge 1$. In other words, $ax^2 + bxy + cy^2$ is almost reduced precisely when τ lies in the famous shaded region pictured (boundary included) in Figure A1.

The exceptions mentioned come from the boundary. The transformation $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ changes τ to $\tau + 1$ relating two points on the vertical boundaries. The transformation $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ relates two symmetric points τ and $-1/\tau = -\bar{\tau}$ on the boundary arc.

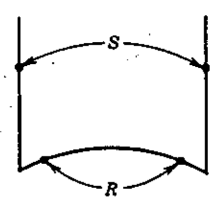

Figure A2

To get rid of the redundant almost reduced forms we throw away half the boundary. Namely:

Definition. $ax^2 + bxy + cy^2 = a(x + \tau y)(x + \bar{\tau} y)$ is *reduced* if τ lies in the region pictured in Figure A3:

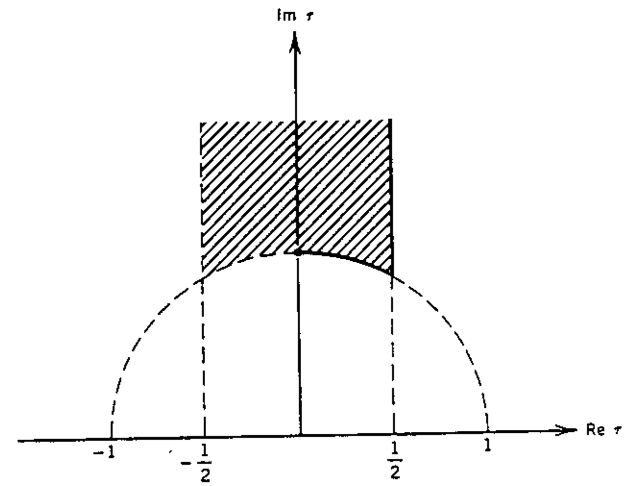

Figure A3

Equivalently, if $|b| \le a \le c$ and in case $a = |b|$, then $b = a$ and in case $a = c$, then $b \ge 0$.

This definition has been made just so that there is a *unique* reduced form in each $\mathbf{SL}_2(\mathbf{Z})$ equivalence class. Hence $\underline{h}(\Delta)$ is the number of reduced forms with discriminant Δ. This leads to a procedure for calculating $\underline{h}(\Delta)$ for a given Δ, namely listing all reduced forms as in Table A1. (The proof of the finiteness of $\underline{h}(\Delta)$ given above shows how to make this list.)

Notice that the forms $2(x^2 + xy + y^2)$ and $2(x^2 + y^2)$ of discriminants -12 and -16 are multiples of forms that appear earlier in the table under $\Delta = -3, -4$. To avoid this multiple listing we modify the game. Define a form $ax^2 + bxy + cy^2$ to be *primitive* if a, b, and c have no common factor greater than 1, and define $h(\Delta)$, the *class number* of Δ, to be the number of *primitive* reduced forms of discriminant Δ. It was a remarkable discovery of Gauss that the set C_Δ of primitive reduced forms of discriminant Δ is an abelian group in a natural way, but we shall not go into that here.*

*Call R_Δ the ring $\mathbf{Z}[\frac{1}{2}\sqrt{\Delta}]$ if $\Delta \equiv 0 \pmod 4$ and the ring $\mathbf{Z}[(1 + \sqrt{\Delta})/2]$ if $\Delta \equiv 1 \pmod 4$. Then C_Δ is isomorphic with the "class group" $\operatorname{Pic}(R_\Delta)$ of R_Δ. When Δ is a fundamental discriminant, then R_Δ is the ring of integers of the quadratic field $\mathbf{Q}(\sqrt{\Delta})$, and $h(\Delta)$ is the class number of that field.

TABLE A1

Δ	$h(\Delta)$	Reduced Forms of Discriminant Δ		
-3	1	$x^2 + xy + y^2$		
-4	1	$x^2 + y^2$		
-7	1	$x^2 + xy + 2y^2$		
-8	1	$x^2 + 2y^2$		
-11	1	$x^2 + xy + 3y^2$		
-12	2	$x^2 + 3y^2$	$2(x^2 + xy + y^2)$	
-15	2	$x^2 + xy + 4y^2$	$2x^2 + xy + 2y^2$	
-16	2	$x^2 + 4y^2$	$2(x^2 + y^2)$	
-19	1	$x^2 + xy + 5y^2$		
-20	2	$x^2 + 5y^2$	$2x^2 + 2xy + 3y^2$	
-23	3	$x^2 + xy + 6y^2$	$2x^2 - xy + 3y^2$	$2x^2 + xy + 3y^2$

TABLE A2

Δ	-3	-4	-7	-8	-11	-12	-15	-16	-19	-20
$h(\Delta)$	1	1	1	1	1	1	2	1	1	2
Δ	-23	-31	-43	-47	-59	-67	-71	-79	-83	-163
$h(\Delta)$	3	3	1	5	3	1	7	5	3	1

With computer assistance these tables have now been extended into the millions.

Looking at the tables one finds that the values $h(\Delta)$ are very irregular, but that with large $|\Delta|$, $h(\Delta)$ tends to be large as well. It has been a fundamental problem to make this last observation precise.

For technical reasons we restrict our consideration for the rest of this talk to the so-called "fundamental discriminants." A discriminant Δ is *fundamental* if it *cannot* be written $\Delta = \Delta_0 f^2$ with Δ_0 a discriminant (i.e., congruent to 0 or 1 mod 4) and f an integer greater than 1. For instance, -12 and -16 are not fundamental. This restriction is not serious because it is known how to compute all $h(\Delta)$ from the values for fundamental discriminants Δ alone.

The fundamental discriminants $\Delta < 0$ with class number $h(\Delta) = 1$ are especially interesting: They are those for which our original problem (find the quadratic equations with a given Δ) has an essentially *unique* solution. One finds easily 9 of them: $\Delta = -3, -4, -7, -8, -11, -19, -43, -67, -163$. Around 1800, Gauss conjectured that there are no more. As we shall see, this is true (but it took more than 150 years to prove).

These discriminants Δ with $h(\Delta) = 1$ have remarkable properties. Let me illustrate with the case $\Delta = -163$.

In 1772, Euler (*Mémoires de l'Académie de Berlin*, extrait d'une lettre a M. Bernoulli) discovered a curious property of the polynomial

$$x^2 + x + 41 \quad \text{(with discriminant } \Delta = -163\text{)}.$$

Namely, if you look at the table of its values for $x = 0, 1, \ldots,$

x	0	1	2	3	4	5	6	7	...	39
$x^2 + x + 41$	41	43	47	53	61	71	83	97	...	1601

you find only *prime numbers*, up to $x = 39$ (but $x = 40$ fails, since $40^2 + 40 + 41 = 41^2$)! The fact that this polynomial yields so many primes is *equivalent* to the equality $h(-163) = 1$. Indeed the following theorem is not hard to prove, using elementary properties of imaginary quadratic fields:

Theorem. For a prime number p that is greater than 3 and congruent to 3 mod 4, the following three properties are equivalent:

a. $h(-p) = 1$.
b. $x^2 + x + (p+1)/4$ is a prime number for every integer x such that $0 \le x \le (p-7)/4$.
c. $x^2 + x + (p+1)/4$ is prime for $0 \le x < (\sqrt{p/3} - 1)/2$.

(For a proof of the equivalence (b) and (c), see, e.g., G. Frobenius, *Gesammelte Abhandlungen* III, no. 94.)

This applies to $p = 163$: By (c), it suffices to check that $x^2 + x + 41$ is prime for $x = 0, 1, 2, 3$; this *implies* it will be so up to $x = 39$.

There are other interesting facts about 163 that are related to $h(-163) = 1$. Consider for instance the transcendental number

$$e^{\pi\sqrt{163}} = 262537412640768743.99999999999925007\ldots.$$

That it is so close to being an integer can be proved a priori from $h(-163) = 1$!

[*Sketch of Proof.* One computes the value of the elliptic modular function $j(z)$ for $z = (1 + i\sqrt{163})/2$; using $h(-163) = 1$, one proves that $j(z)$ is an ordinary integer. On the other hand, the power series expansion for $j(z)$ gives:

$$j(z) = e^{-2\pi i z} + 744 + 196884 e^{2\pi i z} + \cdots$$
$$= -e^{\pi\sqrt{163}} + 744 - 196884 e^{-\pi\sqrt{163}} + \cdots,$$

an expression in which all terms but the first two give a very small contribution (less than 10^{-12}). Hence $e^{\pi\sqrt{163}}$ is close to an integer.]

For these and other reasons, there is great interest in determining all negative fundamental discriminants Δ with class number $h(\Delta) = 1$ (or 2 or 3 or ...).

In the remainder of the talk I will review the work that has been done on this problem, some of it quite recent, some of it still in progress.

The tables suggest that the class number $h(\Delta)$ is roughly of the order of magnitude of $|\Delta|^{1/2}$. One can in fact prove readily that $h(\Delta) < 3|\Delta|^{1/2}\log|\Delta|$.

But we really want a *lower* bound for h, since we want to show that for large discriminants Δ, $h(\Delta)$ must be large as well.

Work of Gronwall in 1913 and Landau in 1918 showed that if the zeta function of $\mathbf{Q}(\sqrt{\Delta})$ has no zero between $\frac{1}{2}$ and 1, then $h(\Delta) > C|\Delta|^{1/2}/\log|\Delta|$ for a constant C which can in principle be computed. Unfortunately, the hypothesis on the zeta function has never been proved (it is a special case of GRH, the Generalized Riemann Hypothesis).

In 1934, Heilbronn completed some previous work of Deuring and proved that $\lim h(\Delta) = \infty$ when $\Delta \to -\infty$. This was soon sharpened by Siegel (1936), who showed that for every $\epsilon > 0$, there exists a positive constant C_ϵ such that $h(\Delta) \geq C_\epsilon|\Delta|^{1/2-\epsilon}$. In other words, the growth rate of $h(\Delta)$ is exactly as expected.

However, Siegel's proof gives less than might be hoped for: It is not "effective" (in plain English, the constant C_ϵ cannot be computed). The reason for this is interesting. One would like to prove that if a discriminant Δ is very large,* then $h(\Delta)$ cannot be too small. One does not know how to do that. What Siegel's proof shows, instead, is that the existence of *two* large discriminants Δ and Δ' with both $h(\Delta)$ and $h(\Delta')$ suitably small leads to a contradiction. This allows $h(\Delta)$ to be small for *one* large Δ, which is one too many!

For instance, it follows from Siegel's work that there is *at most one* fundamental discriminant Δ_{10} with class number 1 beyond the 9 previously listed as already known to Gauss. The question of the existence of Δ_{10} attained notoriety as the "problem of the tenth imaginary quadratic field."

The next progress came in 1952 when Heegner published a proof that Δ_{10} *does not exist*. However, this proof used properties of modular functions that he stated without enough justification. People could not understand his work and did not believe it (I tried myself once to follow his arguments, but got nowhere...). Hence, the question of the existence of Δ_{10} was still considered open.

*I call a negative discriminant "large" when its absolute value is large.

In 1966, Stark studied Δ_{10} in his thesis, and proved that, if it exists, it is very large: $|\Delta_{10}| > 10^{9000000}$. The following year, he succeeded in proving that Δ_{10} does not exist, thus settling the class number 1 problem. His method looked at first quite different from Heegner's; it turned out later that the two methods are closely related (and that Heegner's approach was basically correct, after all).

The same year, A. Baker also gave a solution of the class number 1 problem, by using his effective bounds for linear forms in logarithms of algebraic numbers.

With some work (by Baker himself and by Stark and Montgomery-Weinberger), this method could also be applied to $h(\Delta) = 2$, and yielded the fact that there are exactly 18 negative fundamental discriminants of class number 2, the largest being -427.

However, neither Stark's method nor Baker's applied to the problem of class number 3 or more.

To go further, we must now introduce some new objects. Recall that an *elliptic curve* E over $\mathbf{Q}$ is a nonsingular cubic

$$y^2 = x^3 + ax + b, \quad \text{with } a, b \in \mathbf{Q} \text{ and } 4a^3 + 27b^2 \neq 0.$$

To such a curve is attached a wonderful (and mysterious) analytic function $L_E(s)$, which is called its L series; it is conjectured to extend analytically to the whole $\mathbf{C}$ plane, to have a functional equation similar to the one of the Riemann zeta function (but with respect to $s \mapsto 2 - s$), etc.

This seems to have nothing to do with $h(\Delta)$. However, in 1976, Goldfeld made a startling discovery. He proved that the existence of a *single* elliptic curve E over $\mathbf{Q}$ for which $L_E(s)$ satisfies the preceding conjectures and has a zero at $s = 1$ with multiplicity at least 3 implies

$$h(\Delta) \geq C_E \log|\Delta|$$

for all* Δs, with a positive C_E that is effectively computable. (How can a hypothesis on some elliptic curve imply anything about $h(\Delta)$? Well, it is one of the many mysteries of number theory....)

Goldfeld's theorem tells us that *if* we can find an elliptic curve E with the required properties, then $h(\Delta)$ goes to infinity effectively as $\Delta \to -\infty$. There remains the task of finding such a curve.

There are some elliptic curves, derived from modular forms and called "Weil curves," for which the holomorphy of the L series and the functional equation are known. If we choose for E such a curve, the only further

*This is correct only when $h(\Delta)$ is odd; the general statement is slightly different, see, e.g., [1].

property that is needed is that $L_E(s)$ vanish at $s = 1$ with multiplicity 3 or more. The "Birch and Swinnerton–Dyer conjecture" predicts when this should happen, namely, when the rank of the group $E(\mathbf{Q})$ of rational points of E is ≥ 3. It is easy to find such curves E. One then has to prove

$$L_E(1) = 0, \qquad L_E'(1) = 0, \qquad L_E''(1) = 0.$$

Using the functional equation of L_E (which can be fixed to have a minus sign), this reduces to proving that $L_E'(1) = 0$. But how does one show this? Of course, a computer can check that

$$L_E'(1) = 0.0000000000\ldots$$

accurate to say 10 decimal places. But that is not good enough: The theorem requires $L_E'(1)$ to be *exactly* 0.

No way around that difficulty was found for about 7 years, and as a consequence, Goldfeld's method could not be applied.

The next progress came in 1983, when Gross and Zagier found a closed formula for $L_E'(1)$. Using it, they were able to find a Weil curve E satisfying all of Goldfeld's hypotheses. The corresponding constant C_E has been computed by Oesterlé, and found to be equal to $1/7000$.

To see concretely what this means, let us apply it to the problem of determining the Δs with $h(\Delta) = 3$. Goldfeld's bound gives $|\Delta| \leq e^{21000} < 10^{9200}$. We are thus left with only a finite set of Δs to investigate. Unfortunately, that set is too large.

If the bound 10^{9200} could be brought down to 10^{2500}, one could apply a result of Montgomery–Weinberger saying that, in that range, the largest negative Δ with $h(\Delta) = 3$ is $\Delta = -907$. (Extending the Montgomery–Weinberger method is certainly possible, but would require a lot of computer work.)

Luckily, there are better elliptic curves than the one used by Gross–Zagier. Recently,* Mestre has investigated the rank 3 curve

$$y^2 + y = x^3 - 7x + 6.$$

He has been able to show that it is a Weil curve (this required computer work, too; see a recent note of his, *Comptes Rendus de l'Académie des Sciences*), and, by using the Gross–Zagier theorem, that its L series has a triple zero at $s = 1$. The corresponding C_E turns out to be $\geq 1/55$. For $h(\Delta) = 3$, this gives

$$|\Delta| \leq e^{165} < 10^{72},$$

*This work of Mestre was completed shortly after my Singapore lecture (February 1985).

which is much below Montgomery–Weinberger's 10^{2500}. The class number 3 problem is thus solved. No doubt the same method will work for other small class numbers, up to 100, say.

Of course this is not the end of the story. We would like to have effective lower bounds for $h(\Delta)$ of the size of some power of $|\Delta|$, rather than in $\log|\Delta|$. But how to get them? Will we have to wait until GRH is proved? It may take a while. . . .

References

1. Oesterlé, J., Nombres de classes des corps quadratiques imaginaires, Séminaire Nicolas Bourbaki 1983–84, *Astérisque* **121–122**, Exposé 631.
2. Zagier, D., L series of elliptic curves, the Birch–Swinnerton–Dyer conjecture, and the class number problem of Gauss, *Notices of the American Mathematical Society* **31** 739–743 (1984).